ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 6, Issue 2, pp: (80-84), Month: April - June 2018, Available at: www.researchpublish.com

A Perspective Significant Breakthrough for Machine Translation from Semantic/Syntactic/Episodic Neurolinguistics

Zi-Jian Cai

CaiFortune Consulting, No. 129, Building 6, Room 404, North Dongwu Road, Suzhou City, Jiangsu Province, 215128 PR China

Abstract: **Recently, it was proposed by Cai a new semantic/syntactic/episodic neural model of language, newly encompassing the sentential meanings to the linguistic processes, while deriving three feasible principles from this model to direct machine translation respectively. First, it was necessary to establish the dictionary for translation of words/phrases. Second, it was necessary to read out and translate the grammar of language. Third, it was necessary to determine one correct meaning of some words of multiple meanings by matching them with the episodic/statistical associations with others. Later, it was even shown possible to add some new types of episodic associations for application of this linguistic model to machine translation, such as: (1) Classifying the living/ natural words and phrases by behavior, adopting both the zoological/organizational/physical and affective/ behavioral/logic/characteristic/changing characters to classify the nouns and verbs, helpful to discerning the episodic associations with other words within the sentence or clause. (2) Classifying the sentence/paragraph into the category of natural/social subjects like physics, biology, art, economy, and so on. (3) Collecting the frequent episodic word-pairs for machine translation. Meanwhile, it was recommended to use the episodic symbolization to mark these characters of words in computer following the corresponding words, such as ⊥,<>, and so on. It is a perspective significant breakthrough from neurolinguistics to machine translation.**

Keywords: **Language, Semantic/syntactic/episodic neurolinguistic model, Semantic dictionary, Grammar, Episodic association, Episodic symbolization of word, Machine translation.**

I. INTRODUCTION

Native language is the high and complex mental activity unique to humans for communication and meditation. Recently, the author extended the present declarative/procedural model underlying two kinds of neural processes to a new semantic/syntactic/episodic model of three kinds of neural processes, encompassing the sentential meanings to the neural linguistic processes[1,2].

One of the most important applications for the neural processes of language lies in the promotion of artificial intelligence to assimilate or improve linguistic processes in computer. Machine translation is the most relevant artificial linguistic automation prospectively utilized widely in various situations. Whereas, the results of present machine translation have been quite unsatisfactory, and translational mistakes occur everywhere[3,4,5,6,7], less competent to that of human persons with professional training in translation.

With the recent extension of declarative/procedural model of two neurolinguistic processes to the semantic/syntactic/ episodic model of three neurolinguistic processes[1,2], it is possible to improve the machine translation with the three neurolinguistic processes of brain instead of two, which may bring about a new breakthrough for machine translation. In this article, it is attempted to briefly review these prospective applications.

II. THE BRIEF INTRODUCTION OF THE SEMANTIC/SYNTACTIC/EPISODIC NEURAL MODEL OF LANGUAGE

Recently, Cai proposed a new semantic/syntactic/episodic model of three kinds of neural processes, able to encompass the sentential meanings to the neural linguistic processes[1,2].

Pinker, Ullman and so on integrated a declarative/procedural model for linguistic processing in brain. They suggested that the linguistic lexicon of word-specific knowledge should subserve the storage of meanings of words and phrases, and depend on the temporal-lobe substrates of declarative memory for the storage and usage of facts and events[1,2,8,9,10], while the linguistic grammar should subserve the sequential combination of lexical items by procedural rules into complex sentences, and depend on a network of specific frontal, parietal, basal-ganglia and cerebellar structures of procedural memory in the brain[1,2,8,9,10].

To extend this declarative/procedural model to a new semantic/syntactic/episodic model of three neurolinguistic processes, Cai added that, besides the declarative/procedural neural processes of language, the episodic coordination of cortical modality by reticular formations and thalamic intralaminar projections was also required for language performance and organization[1,2]. Many neuropsychological and pathological evidences supported this supplement[1,2], as in the followings: (1) The reticular dopaminergic system might be involved in the linguistic modality disorganization, since the dopaminergic antagonists[1,2] alleviated while the dopaminergic genes[1,2] phenotyped the stuttering. (2) From drugs assisting the therapies of aphasia, it was implicated that the ascending reticular cholinergic system, noradrenergic projections and nonspecific intralaminar thalamic projections might participate in episodic coordination of cortical modalities in linguistic processes[1,2]. (3) It was evidenced by numerous reports that the gamma bands of high frequency, subject to modulation by the reticular cholinergic system, associated with the word congruency at sentential level[1,2] as well as the episodic coordination of cortical modalities. (4) The N400 from semantic manipulations should be related to the reticular dopaminergic system as it was abnormal in schizophrenic patients when compared to the controls[2], while the P600 of language corresponding to the P3 wave should be related to the reticular noradrenergic system[2].

III. THREE DERIVED PRINCIPLES FOR MACHINE TRANSLATION

Due to the lack of scientific guidance from neurolinguistics, the results of present machine translation are quite unsatisfactory, and translational mistakes occur everywhere[3,4,5,6,7].

With the recent semantic/syntactic/episodic model of three neurolinguistic processes proposed by Cai[1,2], it is possible to derive three corresponding principles from them for guidance of software design of translation machines[3].

From word comprehension as remote declarative memory associations in brain, it could be formulated the first principle for design of translation machines as followings:

Principle 1: Corresponding to the word comprehension as memory associations, it is necessary to establish the dictionary for translation of words and phrases in translation machine.

From grammatical rule as procedural memory in brain, it could be formulated the second principle for design of translation machines as followings:

Principle 2: Corresponding to the grammatical rule as procedural memory, it is necessary to install the linguistic grammar for reading and forming sentence from words and phrases in translation machine, arranging such parts of speech as noun, verb and adjective into order.

From episodic coordination of language in brain, it could be formulated the third principle for design of translation machines as followings:

Principle 3: Corresponding to the episodic coordination of language, it is necessary to work out the episodic associations and statistical concurrence for the words and phrases of multiple meanings, and figure out one correct meaning of them in translation machine.

The three principles for guidance of software design of translation machines are briefly outlined in Table 1.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 6, Issue 2, pp: (80-84), Month: April - June 2018, Available at: www.researchpublish.com

Table 1: Three principles from brain neurolinguistics for machine translation

	Neurolinguistic Processes in Brain	Principles of Machine Translation
Principle 1	Semantics as memory associations	Make dictionary of words/phrases
Principle 2	Syntax as procedural memory	Install linguistic grammar
Principle 3	Episodic coordination of language	Work out the episodic and statistical word meanings

IV. EPISODIC ASSOCIATIONS FOR APPLICATION OF THIS LINGUISTIC MODEL TO MACHINE TRANSLATION

To determine the meaning of a word/phrase of multiple meanings in the sentence or clause, the episodic association of the word/phrase with other words in the sentence or clause can provide some important clues, because its meaning varies in episodic association with other words in the sentence or clause.

It would be useful to formulate some common episodic methods for identification of the word/phrase of multiple meanings in the sentence or clause. Herein it is briefly reviewed some of the methods[11], as the followings:

(1)The author suggested that most words could similarly be classified according to the characteristics of them to possibly manifest as a living entity or a natural matter, just as the present pronouns “he”, “she” and “it”. Besides, at the individual level, all sentential contents could fall into either physical sciences or life sciences, while at the organizational level, all sentential contents could fall into either natural sciences or social sciences. In this regard, the physical and zoological category contained all nouns at the individual level, while the organizational category contained all nouns at the organizational level[11].

The nouns in the physical and zoological category adopted different verbs to manifest their behaviors respectively. The physical verbs could correspondingly be classified into characteristic, changing and logic characters, while the zoological verbs could be classified as affective, behavioral and logic. Therefore, all nouns and verbs could be classified into both zoological/organizational/physical and affective/behavioral/logic/characteristic/changing characters[11], even able to adopt two separate characters of Vt to associate it with subject and object in sentence. Likewise, the adjectives and adverbs could also be classified into the affective/behavioral/logic/characteristic/changing/spatial/temporal characters[11].

(2) The meanings of words within the sentence mostly fitted the category of subject of the sentence, such as the physics, biology, art, economy, and so on, in the natural sciences or social sciences. Categorization of the sentential subject could certainly help translation, with some words of multiple meanings being determined to one correct meaning fitting the category of such subject[11]. Besides, the people’s representatives in Beijing in television on January 11, 2017, strongly argued that they often deviated the main topic while making speech, so that the thematic subject should be improved to the paragraphic subject for the purpose of translation[11].

(3) In some situations, it is also helpful to refer to the frequent word-pairs to determine the episodic associations. It would thus be useful to collect and store the frequent episodic word-pairs in computer[11].

V. EPISODIC SYMBOLIZATION OF WORDS FOR COMPUTER APPLICATIONS

It is necessary to consider how to implement these applications from this linguistic model to the computer. The words and their episodic associations could be stored in computer as various databanks for SQL or Oracle, while the grammar and translation could be read out by programmed software.

For storing the episodic associations of a word/phrase of multiple meanings in SQL or Oracle, it was recommended to use the symbol ⊥ to separate the multiple meanings[11], to use another symbol <> to characterize the behavioral classification and subject category of each meaning of the word/phrase[11], with the symbols /, - and * within <> to separate the various characters. With this episodic symbolization, a word of multiple meanings could be stored as if one data including all its meanings in SQL or Oracle.

For example, the noun “charge” has several meanings as “duty⊥money-payment⊥payment-amount⊥electrical-quantity”, and could be better stored as one complex data in databank as “duty<Nn-zoological/organizational/behavioral>⊥money-payment<NvNn-zoological/organizational/behavioral>⊥payment-amount<Nn-physical/characteristic>⊥electrical-quantity<Nn-physical/characteristic><p-physics>”.

In translation, all words/phrases of multiple meanings in the sentence or clause are directly transformed from the databank into the complex stored form with ⊥ to separate the multiple meanings. Then, according to the contents in the <>, the computer would make the additional choice via the episodic and statistical associations with others in the sentence or clause to figure out the correct meanings.

It is necessary to point out that there are still some other methods able to improve the translation, such as using the visual and auditory inputs, characters, associations, and so on. This article only reviews the improvement of translation prospectively brought about from the new semantic/syntactic/episodic model of language, while does not consider those from others.

VI. CONCLUSIONS

In this article, from the three neurolinguistic semantic/syntactic/episodic model of Cai, it is reviewed three principles for guidance of software design for translation machines, as followings: (1) Principle 1 as the necessity to establish the dictionary for translation of words/phrases in translation machine. (2) Principle 2 as the necessity to install the linguistic grammar for reading and forming sentence from words/phrases in translation machine. (3) Principle 3 as the necessity to work out the episodic and statistical associations for the words/phrases of multiple meanings and to figure out one correct meaning of them in translation machine.

In this article, it is also reviewed three types of episodic associations prospectively able to improve the machine translation from the semantic/syntactic/episodic model of language, as followings: (1) The behavioral characteristics both as zoological/organizational/physical/categorical and affective/behavioral/logic/characteristic/changing to further classify the living/natural nouns and verbs, the affective/behavioral/logic/characteristic/changing/spatial/temporal characteristics to the living/natural adjectives and adverbs. (2) Classification of the sentence and paragraph into the category of natural/social subjects such as physics, biology, art, economy, and so on, so as to help discern the meaning of words/phrases of multiple meanings with paragraphic subject. (3) Collection of frequent episodic word-pairs in computer, so as to help determine the episodic associations.

Finally, it is briefly recommended the suggestion to use the episodic symbolization of words to apply the above progressions to computer. The words/phrases and their episodic associations could be stored in computer as various databanks for SQL or Oracle, with the symbol ⊥ to separate the multiple meanings, and the symbol <> to characterize the behavioral classification and subject category of each meaning of the word/phrase of multiple meanings, and so on.

Funding:

The author declares no financial support for this work.

Conflict of interest statement:

The author declares no conflict of interest for this work.

ACKNOWLEDGMENTS

It is herein acknowledged that Mingxun Cai paid the Open Access publication fee of this paper.

REFERENCES

[1] Cai ZJ (2015) Semantic memory association, procedural grammar syntax and episodic modality coordination as three interactive neural processes organizing language: A model. OALib J 2: e1718. doi: http://dx.doi.org/10.4236/oalib.1101718.

[2] Cai ZJ (2017) The neurobiological models of language: An updated review. Psychol Neurosci 10(3): 297-306.

[3] Cai ZJ (2016) Principles derived from neurolinguistics of brain for design of translation machines. OALib J 3: e 2704. doi: http://dx.doi.org/10.4236/oalib.1102704.

[4] Zhang M, Duan XY, Chen WL (2014) Bayesian constituent context model for grammar induction. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22(2): 531-541.

[5] Zhang M, Che WX, Zhou GD et al (2008) Semantic role labeling using a grammar-driven convolution tree kernel. IEEE Transactions on Audio, Speech, and Language Processing 16(7): 1315-1329.

[6] Xiong DY, Zhang M, Li HZ (2011) A maximum-entropy segmentation model for statistical machine translation. IEEE Transactions on Audio, Speech, and Language Processing 19(8): 2494 – 2505.

[7] Xiong DY, Zhang M, Wang X et al (2015) Topic-based coherence modeling for statistical machine translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing 23(3): 483 - 493.

[8] Pinker S (1991) Rules of language. Science 253: 530-535.

[9] Ullman MT, Corkin S, Coppola M et al (1997) A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. J Cogn Neurosci 9: 266-276.

[10] Ullman MT (2004) Contributions of memory circuits to language: the declarative/procedural model. Cognition 92: 231-270.

[11] Cai ZJ (2017) Three types of episodic associations for the semantic/syntactic/episodic model of language prospective in applications to the statistical translation. OALib J 4: e3830. doi: http://dx.doi.org/10.4236/oalib.1103830.